INSTRUCTOR NOTES FOR M358K FOR PART III (GATHERING DATA: CHAPTERS 11 - 13) OF DEVEAUX, VELLEMAN AND BOCK, *STATS: DATA AND MODELS*, 3RD EDITION

Chapter 11: Understanding Randomness (pp. 267 – 280)

- *General Comment/*Caution: This chapter seems to be good for what it is trying to do, but I'm not sure that what it's trying to do is in all cases what's best for the course. So I recommend reading the comments below and the whole chapter before deciding if you want to follow what is here or redo it to better suit what you think is best for the course.
- pp. 267 – 269 I have no problem with this part. It is a good introduction which brings up some points where students might have intuitions that need reexamination. (Note: The blue box on p. 269 may be important for some international students.)
    - You might want to do the "pick a number" exercise in class before having the students read this chapter.
- pp. 269 - 275 Here's where I wonder if some things might have been done better differently (although some things here are done very well).
    - The examples used here seem pretty removed from the rest of the subject matter of the course.
    - On the other hand, the first example (pp. 269 – 270) serves well to introduce the idea of a random model and simulation, and the vocabulary of "trial" and (less standard) "component."
    - Still, our students have encountered the idea of "trial" in M362K.
    - Returning to the other hand, the explicit emphasis on "Specify how to model a component outcome …" ,"Specify how to simulate trials," "Put it all together …" and "Analyze the response variable" (pp. 270 – 271) is good.
    - But do we want to have our students doing simulations involving random digits? I'm not entirely convinced that this is still appropriate.
    - Hand 2 again: The Just Checking (p. 272), Step-by-Step Example (pp 273-274), What Can Go Wrong (p. 274) and Connections (p. 275) make some good points.
- I'm also not convinced that this chapter gives an adequate introduction for the type of computer simulations that we may want to use.
- In case you are considering seriously replacing this chapter, here are places I have found later in the text that refer explicitly to this chapter:
    - p. 287 uses random numbers to obtain a simple random sample.
    - p. 337 refers to "trial" (but essentially defines what the term means)
    - p. 339 uses random numbers to simulate flips of a coin.
    - pp 404 – 405 use the sports card context to introduce Bernoulli trials and the geometric model – the latter of which, as mentioned below under Chapter 17, is not very relevant to the material in this course (except for not confusing it with the binomial model)

  - pp. 439 – 440 uses simulations, but not in a way that uses the specifics of this chapter
  - p. 447 mentions simulation, but in a way that really doesn't refer to what's in this chapter.
- The general comments in the "Simulation on the Computer" (p. 276) are worth assigning as reading.
- Suggested exercises (pp. 277 – 280): Some of 1-4 and 9-10 for class discussion; #5, 7, 11 for self-check; #12 written.

Note re chapters 12 and 13:

- You may want to pass out the Term Project Description for students to read while reading chapters 12 and 13. (See Comments for Instructors on Projects and Project Handouts; both are linked from http://www.ma.utexas.edu/users/mks/M358KInstr/M358KInstructorMaterials.html)
- Chapters 12 and 13 especially lend themselves to assigning reading and exercises, then devoting class time to questions, discussion, and examples rather than lecturing.

Chapter 12: Sample Surveys (pp. 282 – 305)

- p. 285
  - Note the references (under Populations and Parameters) to Chapters 6-8. You will need to explain these if you have decided to cover Part III before these chapters.
  - The activity Learn About Statistics and Parameters from the ActivStats disk (as best I can figure, this is the one referred to as Statistics and Parameters on p. 285) gives a nice preview of statistical inference. I recommend at least looking at it yourself, then either showing it in class or giving a short lecture yourself along similar lines. One place to do this would be in connection with the section Populations and Parameters (pp. 285 – 286). Another place might be at the end of Chapter 13. Or you might want to divide the content into two short lectures, one at each of these places.

- p.286
  - Note that the book's definition of "representative" is different from the ordinary/intuitive use.
  - The Just Checking is good.
  - The statement "An SRS is … the sampling method on which the theory of working with sampled data is based," may be subject to misinterpretation. It should *not* be interpreted as saying that all inference techniques require SRS's.
- p. 287 An example illustrating sampling variability might be helpful in making the point. One of many possibilities: Assuming packages of candies such as M and M's are filled in a process that assures random mixing, the contents of individual packages would constitute simple random samples (more or less). It is easy to see (if not common experience) that individual packages of the same size

often have different numbers of each color. (When after-Halloween sales come along, you might want to invest in some packages of small-size packets of some such candy, to use in a class activity where each person uses their small packet as a sample to estimate the proportion of reds or whatever someone's favorite color is. Or large packages, doling out a scoop to each student, also work.)

- pp. 291 – 292 Point out that projects will require more analysis (to be covered in later chapters) than in this example.
- p. 292 The point in the blue box is important!
- pp. 293 – 294
    - There are whole books in the subject of phrasing survey questions carefully. For references and further examples, see http://www.ma.utexas.edu/users/mks/statmistakes/wordingquestions.html
    - If students want to do a project involving asking people questions, strongly urge them to do a pilot study (as mentioned at the bottom of p. 294) to help figure out good wording. A pilot study – conducted before writing a project proposal -- is often a good idea for other projects as well, especially experiments.
- pp. 296 - 267 The "What Can Go Wrong" section is especially relevant for students planning to do a project involving a sample survey – or if you allow students to do a project using existing data.
- p. 298 The "Connections" box is unusually good.
- p. 300 "Sampling on the Computer" may be worth assigning as reading.
- The String Sampling Activity (See http://www.ma.utexas.edu/users/mks/M358KInstr/M358KInstructorMaterials.html) is a worthwhile addition to Chapter 12. It seems to be effective in helping students learn and remember that a sample might be biased even though the sampling process appears random.
- Sampling bias is sometimes called *ascertainment bias*, particularly in biology. It has occurred in studies of human genetic variation, which typically use DNA microchips to identify variation in certain genes that are known to have different versions. If the microchip is created to assess only certain genes known to vary in a particular population, the study will not pick up genes that do not vary in that population but vary between that population and others, or within some other populations. For example, a study using a microchip based on genes known to vary in European populations may miss variation between European and Asian populations or between different Asian populations. (See Box 1, p. 600 in Jobling, MA and Tyler-Smith C (2003), The human Y chromosome: an evolutionary marker comes of age, *Nature Review Genetics*, 4(8):598-612)
- Exercises (p. 301 – 304): I'd suggest odds from 1 – 9 for self check and class discussion (and/or perhaps evens from 2 – 10 for class discussion), plus a selection from 22, 24, 26, and 30 to hand in.

Chapter 13: Experiments and Observational Studies (pp. 305 – 329)

- This chapter seems particularly well done.
- pp. 308 – 310 Generally well done. Some suggested additions:

  - p. 308 Some students may find it interesting to watch the video "An Industrial Experiment" on the disk included with the text.
  - p. 309 (item 3) Add the message: "As we'll see in Chapter 21, replicates under the same conditions are also important." (In other words, the possibility of Type I error means we can't claim something as fact just because we get a statistically significant result in one experiment. That's why the FDA often requires two clinical trials before approving a new drug. You and some of your students might find the recent article, Ed Yong, Replication studies: Bad copy, *Nature*, 17 May 2012, doi: 10.1038/485295a, http://www.nature.com/news/replication-studies-bad-copy-1.10634 interesting.)
  - p. 309 (Item 4) Add the message: "Experiments involving blocking require a different method of analysis than those not using blocking. This is discussed for some blocking designs in Chapter 25."
- pp. 310 – 311 Aspects of the step-by-step example are important for proposals for projects involving experiments. (However, note that the experiment outlined here requires ANOVA, which is not a required part of the syllabus for M 358K)
- pp.312-313 This seems like a good "looking ahead" to inference.
- pp. 313 – 314 The sections "Experiments and Samples" and "Control Treatments" are especially well done.
- p 319 Be sure to emphasize the last sentence before the section Confounding: "You can take a whole course on the design and analysis of such experiments." Maybe go further and caution that this course just gives an introduction to statistics – there are *many* situations that require more sophisticated analysis methods than in this textbook.
- p.322 The Connections section seems especially good and is worth emphasizing.
- Exercises (pp. 325 – 330): I suggest 1, 3, 5, 9, 13, 21, 29, 31 33 for self-check and/or class discussion, plus perhaps a couple of evens to hand in.

Review Exercises (pp. 331 – 335): I suggest #1, 3, 9, 25, 31, 41 for self-check/class discussion, and perhaps 4, 14, 28, 42 to hand in.